USE OF PARALLEL COMPUTING IN MASS PROCESSING OF LASER DATA

ZASTOSOWANIE OBLICZEŃ RÓWNOLEGŁYCH DO MASOWEGO PRZETWARZANIA DANYCH LASEROWYCH

Janusz Będkowski, Ryszard Bratuś, Marcin Prochaska, Antoni Rzonca

DEPHOS Software sp. z o.o., Zabierzów

KEYWORDS: parallel computing, laser scanning, lidar data processing

SUMMARY: The first part of the paper includes a description of the rules used to generate the algorithm needed for the purpose of parallel computing and also discusses the origins of the idea of research on the use of graphics processors in large scale processing of laser scanning data. The next part of the paper includes the results of an efficiency assessment performed for an array of different processing options, all of which were substantially accelerated with parallel computing. The processing options were divided into the generation of orthophotos using point clouds, coloring of point clouds, transformations, and the generation of a regular grid, as well as advanced processes such as the detection of planes and edges, point cloud classification, and the analysis of data for the purpose of quality control. Most algorithms had to be formulated from scratch in the context of the requirements of parallel computing. A few of the algorithms were based on existing technology developed by the Dephos Software Company and then adapted to parallel computing in the course of this research study. Processing time was determined for each process employed for a typical quantity of data processed, which helped confirm the high efficiency of the solutions proposed and the applicability of parallel computing to the processing of laser scanning data. The high efficiency of parallel computing yields new opportunities in the creation and organization of processing methods for laser scanning data.

1. INTRODUCTION

This paper presents selected results produced by a 3-year research project financed by the European Union and completed in the period 2012 - 2015 by Dephos Software Ltd in Kraków (Poland). The title of the project was: "Research on large scale storage, sharing, and processing of spatial laser data." The project covered two basic, overlapping areas of research – the second of which (large scale processing of laser scanning data) is discussed in this paper.

The initial idea for the project stems from the practical experience of people working for the Dephos Software Company in Kraków. In 2004 the Dephos photogrammetric station became equipped with additional capabilities designed to support terrestrial and aerial laser scanning technologies. The company was already broadly employing such pioneering technologies in its production department at the time (Rzonca, 2006). Research and development work at the time focused on the creation of an effective format for recording point cloud pyramids, custom systems to yield an absolute orientation of scans as well as to display and vectorize scans, and the still unique capability of stereoscopic observation of point clouds. A number of new algorithms designed to generate so-called orthoscans were created using custom technical knowledge. In addition, tools designed to color point clouds were created along with tools designed to help "fly" through point clouds. Early attempts were also made to yield a new classification system based on the analysis of terrain geometry via cross sections.

The project was made to research and implement contemporary parallel computing methods and the experience of programmers and photogrammetrists working with Dephos. New research and development was needed, as new computing methods also affected the algorithms used. The custom algorithms and implementation modification were made basing on modern GPGPU (General Purpose Graphical Processing Unit) processors application, which can be used in spatial data processing (Będkowski *et al.*, 2013).

In light of the research literature, self-collected knowledge of photogrammetry and computer science, the following hypothesis may be formulated: The use of parallel processing using a graphical processor with custom algorithms for the processing of laser data obtained from the DEPHOS Software Company will yield a major improvement in process parameters. This is especially true of processing time, output quality, and the simplification of steps needed to prepare data for processing. The investigated algorithms varied in terms of degree of complexity and level of innovation. Some were modified to be used exclusively with graphical processors. Others represent a novel approach to both processing and laser scanning in general. The publication provides the results of research and discusses various aspects of the solutions used.

2. LIDAR DATA AND PARALLEL PROCESSING

Lidar data, also known as laser scanning data, feature a number of special characteristics. Data of this type consist of point clouds identified by three coordinates as well as additional data – often more than ten types according to LAS specifications.

(http://www.asprs.org/a/society/committees/standards/asprs las format v12.pdf)

There are millions of points in a single scan. This is easier to imagine if we consider an informal efficiency criterion for displaying point clouds to be 100 million points. If a graphical user interface possesses the ability to display simultaneously such a large number of points, then it can be described as fully effective. The experience of people working for Dephos and Dephos Software has been that such a large number of data points is difficult to display, store, and process. In the case of R&D work pursued by programmers at Dephos Software in the area of laser scanning technology, such large numbers of points require a special approach at the very least.

Research on scanning algorithms at Dephos was accompanied by the emergence of parallel computing based on graphics card processors made by nVidia. (http://www.nvidia.pl/object/cuda-parallel-computing-pl.html) The CUDA cards made by nVidia appeared to be the perfect solution in light of the nature of scanning data (Han *et al.*, 2009; Zeng, He, 2009)). The requirements of this new technology can be summarized to the following: Computational processes can be programmed and completed in parallel when

exists N simple and analogous processes – which can be run independently on a larger number of data, and results can be obtained in a very similar amount of time.

Theoretically, there exists an opportunity to accelerate selected processes (e.g. 1,000 times) relative to serial computations done using advanced multi-core processors. A large part of the research focused on matching the nature of scanning data and resulting consequences and needs as well as the capabilities of CUDA graphics cards. Therefore it became necessary to test the degree to which process acceleration is possible when algorithms are adjusted to the requirements of scanning data.

The following paragraphs discuss selected research results divided into two groups: basic processes and advanced processes.

3. RESEARCH METODOLOGY

The research described herein proceeded in the following manner. A team of specialists in the area of photogrammetry and laser scanning designed algorithms to generate orthoimages or classify point clouds. Such algorithms were then adapted to the calculation methods used with graphical processors.

At the point when an algorithm was found to work properly in a given numerical range, it was employed by a research team working on photogrammetry and scanning. Once results were found to be accurate, efficiency tests were performed along with comparative analysis based on commercially available algorithms used in production. The testing work focused on the comparison of processing times alone for the studied algorithm and analogous commercially available tools. It was then discovered that not all commercially available algorithms yield a credible measurement of processing times. In certain cases, the authors of the paper had to measure time components, i.e. loading, processing, and recording, which naturally made the results less credible.

The process was repeated for a number of data sets with variable processing parameters as well as using a variety of hardware configurations. This publication provides a part of a large amount of research material.

4. ANALYSIS OF SELECTED BASIC PROCESSES

4.1. Generation of ortho-images

The first basic scanning data processing step consists of the generation of ortho-images by calculating RGB components for each pixel situated on a predefined projection plane. For a process to run properly, it is necessary to enter basic parameters such as pixel size, radius of the color search, as well as additional parameters that make it possible to fill gaps between point images on the orthoscan, but without creating artifacts (Pyka, Rzonca, 2006).

The hypothesis that the process of generating ortho-images can be substantially accelerated was produced by a comparison between the description found above and the requirements and limits of the parallel algorithm of interest. Empirical tests performed on large scale data helped to confirm this hypothesis. Parallel processing proved to be efficient to the point where given the assumption that 100% is the time needed to process one million points via the two studied methods (serial – CPU, parallel – GPU), the time needed to open and save files equals 76.5% of the total time. The CPU process uses 23.0% of total

time, while parallel processing time is a mere 0.4% of total time. A similar pattern can be observed for a sample of 11 million points: 86.5% (open and save time), 12.6% (CPU), (0.9% GPU).

Research and development work provided a comparative analysis of a typical production tool such as Bentley PointTools and a new parallel processing tool. Samples consisting of 2, 6, and 32 typical ISOK sections were used in the study. This was the equivalent of about 25 million, 75 million, and 400 million data points. ISOK is the Polish acronym for a national information systems security project operated by Poland's government. Figure 1 shows processing times that confirm the superior quality of parallel processing, particularly for huge number of points. The analysis did not include the time-consuming conversion of data from the LAS to the POD format, which is a required step when using Bentley PointTools.



Fig. 1. Large scale generation times for ortho-scans via the serial and parallel methods.

4.2. Coloring point clouds

The process of a coloring point cloud is a process of integrating photogrammetric data and scanning data (Rzonca, 2013). This is done by computing RGB components for each point and assigning them to that point in a point cloud file. This process is more complex than that described for the generation of orthoscans. The calculation of a color for a given point occurs on the basis of colors extracted from one or more adjacent photographs. This process may be expressed in simple terms as the use of the collinearity equation to find a color for a point in a cloud where the plane of a photograph is intersected by a vector connecting the projection center of the photograph with the point being colored.

A number of experiments were conducted in order to help determine the usefulness of the parallel processing method in the point cloud coloring process. Figure 2 shows the outcome of the first experiment -a comparison of the time needed to open, upload, and save data (i.e. data transfer) relative to the time needed to perform calculations alone using a CUDA card made by nVidia.



Number of points (in millions)

Fig. 2. Relationship between time, data transfer rates, and data coloring rates using CUDA parallel processing technology by nVidia.

The second experiment consisted of a comparison of processing timesneeded when using a commercially available tool such as Terrasolid TerraPhoto and the tool created by Dephos (CuScanColorizer) based on the principle of parallel computing. The amount of time needed to process 11 million points was reduced by 47%. This does not include the large amount of time and effort needed to prepare data when using TerrasolidTerraPhoto. R&D work in the area of point cloud coloring has not yet been finished by this time. It is expected that the quality of the CuScanColorizer tool will increase with a new option that will become available for large scale processing of entire blocks of photographs and point clouds.

4.3. Transformation of cloud coordinates

The third area of basic processing consisted of simple calculations or the transformation of cloud coordinates. Transformation times were compared for different numbers of points for four cases: (1) using a parallel computing , (2) single core serial processor, (3) dual core serial processor, (4) eight-core serial processor. The best results were obtained via parallel processing – as shown in Figure 3.



Fig. 3. Computing time for different quantities of data and different types of transformation – parallel computing (GPU GeForce GTX 970m), multi-core computing (1,2,8 CPU Intel i7-4710HQ @ 2.5GHz).

4.4. Generating a regular grid

The last basic method of processing a point cloud, examined as part of this study, was the generation of a regular grid. Shepard's method was used to generate a regular grid (Franke, Nielson, 1980). This is a mean of multidimensional approximation of dispersed groups of points. Shepard's method introduces a weighted coefficient inversely proportional to the p^{th} power of the distance to a given point. The *p* parameter affects the value of the coefficient. The weighted coefficient w_i for point *x* is calculated via the equation (1).

$$w_i(x) = \frac{1}{d(x,x_i)^p} \tag{1}$$

Results obtained as part of research and development workincluded processing time only in the first experiment, and time to completion– including upload time – in the second experiment (Fig. 4). The measurement of processing time alone using GPU exhibited an approximately 10-fold acceleration relative to a classic algorithm. A sample of about 11 million points was used to generate a grid for four different values of the p parameter (p = 1, 2, 5, 10) via serial processing and parallel processing. The results of the comparison can be observed in Figure 4.



Fig. 4. Grid generation times via serial processing (CPU) and parallel processing (GPU).

In summary, all three basic processes produced substantial acceleration, which confirms the hypothesis that the use of parallel processing with algorithms processing point clouds is an effective strategy that results in substantially faster output. The effectiveness of this method is not in doubt, but the optimization of uploading and saving processes is necessary in order to yield a complete improvement in process efficiency. Perhaps the best solution consists of an custom, flexible saving format and the serial use of multiple processes operating via this format. This proposed solution would eliminate the change in format to the exchange format known as LAS.

5. ADVANCED PARALLEL PROCESSING OF POINT CLOUDS

Advanced point cloud processing methods including the three presented in this paper are complex processes and their output cannot be evaluated using merely a simple comparison of completion times. An evaluation of algorithm correctness ought to involve a broader analysis of a variety of quality-related issues.

The scope of this paper does not allow for a publication of all research results, especially since the purpose of the paper is to provide an overview of parts of a project completed by the Dephos Software Company.

5.1. Plane detection

The first process to be used in research and development in this studywas the process of plane detection. This part of the study involved the use of the RANSAC algorithm (Fischler, Bolles, 1980; Fischler, Bolles, 1987), which was adapted to use with parallel processing. The RANSAC plane detection algorithm was tested on two sample point clouds – one with 1.5 million and the other with 15 million points. The experiment included an examination of the effect of the number of iterations on processing time as well as the number of planes detected. Figure 5 shows the relationship between processing time and the number of iterations (between 10 and 200). The detection time needed to find 49 planes was shown to be less than 2 minutes even for a point cloud with 15 million points (ground-based scanner data).

The results produced by this study confirm the need for the use of parallel processing also with this type of advanced process.



Fig. 5. Relationship between plane detection time and the number of iterations.

5.2. Classification of point clouds

The next advanced process, one that is especially important in the case of the processing of scanning data, is point cloud classification. The most popular tool in scanning production at this time is Terrasolid Terrascan. The new proposed algorithm uses an unique geometric rule as a basis for classifying data points. In addition, a large number of its operations occur in parallel. However, this method cannot be used for all stages of the classification process. The first feature to be recognized is the ground and then buildings and other classes

available via the LAS standard. In this study, these algorithms were evaluated for their efficiency and the results were then compared with those for a cloud classified automatically via the Terrascan tool.

Efficiency tests were conducted on large data samples – a total of 668 ISOK sections (see earlier definition). Process completion times were measured along with loading times and the saving of changes. Despite a lack of optimization in terms of opening and saving in the LAS format, the observed increase in completion time was only about 60% larger (Fig. 6).



Fig. 6. Classification times via serial processing and parallel processing.

The basic difference lies in the quality of processing. A cloud classified using the Dephos method as part of this research study does not require any manual work on the part of the operator. Figures 7, 8, 9 and10 provide examples of properly classified walls of buildings. Figure 7 shows a point cloud classified using a commercial algorithm, while Figure 8 shows the outcome for the algorithm used in this research study. It is important to note that the walls of buildings were properly placed in the right classes of buildings using the novel algorithm, which results are presented below in this paper.



Fig. 7. Class of buildings following classification using Terrasolid Terrascan – walls are not properly classified.

Janusz Będkowski, Ryszard Bratuś, Marcin Prochaska, Antoni Rzonca



Fig. 8. Class of buildings following classification using the Dephos tool – walls are properly classified.



Fig. 9. Cross section of a classification of Terrascan software.



Fig. 10. Cross section of a Dephos classification.

Figures 9 and 10 illustrate a similar pattern. Vertical elements are placed in the right classes without error.

In summary, the studied algorithm is at the testing and implementation stage – and requires optimization, especially in the area of operations outside of classification alone. Not every operation associated with this algorithm can be recorded in parallel form. At the same time, the quality of the classification along with the planned optimization steps constitute a basis for the assertion that the studied DEPHOS algorithm is an algorithm characterized by high technological value – at least at the research and implementation stage.

5.3. Data analysis for the purpose of quality control

An important area in advanced scanning data processing is the analysis of data quality. This is a broad area of discussion, but one that includes the following four basic actions: (1) data cutoff based on scanning angle, (2) classification of points in overlapping areas between

strips (so-called overlaps), (3) control over diagonal coverage, (4) statistical analysis of the distribution of intensity values, uniformity of density, and the last reflection.

The use of parallel computing for the purpose of angle cutoffs has not brought the expected results thus far in the areas of efficiency and classification. Processing times have increased about 70% as a result. The algorithm's accuracy has already been confirmed, but it needs to be optimized to solve the problem of hard disks overloading by huge amount of result data.

A custom approach was used to classify points in overlapping areas. The use of parallel computing has made it possible to consider factors other than the angle of scanning and the boundary of the strip, which is standard practice in TERRASOLID Terrascan software. Thanks to improved computational capabilities, the DEPHOS algorithm also considers the shape of the cloud, number of neighboring points, and preset density, which results in the keeping of points hidden in one strip and visible in another. If only the scanning angle and the boundary of the strip are considered, then such points are unnecessarily removed. Figures 11 and 12 show the results of the classification of points in the overlap area produced using commercially available software (Terrascan) and that produced using DEPHOS software.



Fig. 11. Classification of overlap points using Terrascan software.



Fig. 12. Classification of overlap points using parallel processing.

Despite parallel processing being much more complicated and relying on more stages, the speed of computations increased 10-fold for 668 ISOK sections (see earlier definition).

Two additional stages of data analysis identified in the study were characterized by very high computational efficiency, but this efficiency could not be compared with analogous CPU-based solutions – due to the lack of a commercially available analytical tool. Control over diagonal coverage is currently performed on a fixed number of cross sections with the use of parallel computing thanks to the high efficiency of calculations, which provides an opportunity for accurate analysis. Completion times lasting mere seconds were obtained in

mass processing associated with statistical analyses of the distribution of intensity values as well as the uniformity of density and the last reflection.

The use of parallel computing in data analysis makes it possible to produce faster and more accurate calculations without the need to divide data into sections, as is the case with other software such as Terrascan. The DEPHOS algorithm is efficient enough to work on original data from the flight mission– files corresponding to entire strips. This makes work much more efficient and the amount of time saved is actually difficult to estimate and it was not used in comparative analyses.

6. DISCUSSION OF RESULTS

In most cases, the research results provided herein have confirmed the research hypothesis. The use of parallel processing with graphical processors reduced processing times markedly. It was not always possible to compare processing times alone. However, research suggests that process efficiency can be improved significantly. This is possible in part due to various technological improvements that make it easier to prepare data for processing purposes and ultimately improve output quality.

Significant reductions in processing times were obtained for orthophoto generation and point cloud coloring. In the case of coordinate transformation, shorter times were noted compared with serial processes run on processors with as many as eight cores. An algorithm designed to generate a regular point grid based on a point cloud was run using a variety of processing parameters and exhibited significant acceleration with a GPU processor.

In the case of advanced processes, high efficiency was obtained with a process designed to identify planes in a point cloud. For the classification created using a custom algorithm, the processing time was slightly better than that for an analogous product made by Terrasolid. It was also the case that the former produced fewer classification errors in critical parts of the point cloud. The custom methods described in this paper also produced much better results in the area of processing quality in point cloud analysis. However, processing times for the complete process were worse due to a delay in output recording on the disk. This outcome is not an indicator of the overall efficiency of the process itself, as it includes non-optimized file opening and saving times.

The research results may be summarized in the following manner. The research resulted in a substantial reduction in processing times in the case of basic processes. On the other hand, the results for advanced processes appear promising in the area of processing times and the area of product quality thanks to the use of innovative solutions in point cloud processing.

7. SUMMARY AND CONCLUSIONS

The research executed by DEPHOS Software sp. z o.o. has shown that the high efficiency of parallel computing yields new opportunities in the creation and organization of processes applied to laser scanning data.

A large increase in efficiency was observed in the area of basic processing with respect to the generation of orthoscans, coloring of point clouds, transformation of point clouds, and generation of a regular grid.

In the case of advanced calculations, it proved to be difficult to compare the efficiency of parallel computing with that of algorithms based on serial computing means. In addition to completion time, the quality of output was also considered. The use of innovative methods produced an increase in processing efficiency and product quality in the area of detection of planes and classification of point clouds. In the case of processes part of so-called statistical analysis of point clouds, the following results were obtained: Data cutoff based on angle in the case of parallel computing increased completion time by 70%. This is the only negative result of the research published in this paper. This suggests that loading mechanisms need to be optimized, as do saving mechanisms. On the other hand, the classification of overlap points proved to be ten times more efficient. Next, in the case of the analysis of diagonal coverage and statistical analysis (intensity distribution, uniformity of density, the last reflections), no corresponding software was identified in the marketplace; however, process efficiency was found to be high and the same was true of the quality of the end product.

In summary, in the face of such efficient processes, it is possible to create two processing procedures: (1) sequential processing based on a single loading, operation of several groups of parallel processes – one following the other – and recording of results; as well as series processing consisting of ordered repetitions of a given process using different parameters. The parameters selected for future use are those shown to yield the highest product quality, as shown by automatic or manual quality control efforts.

The research discussed in this paper will be continued in order to implement the findings, as current results are promising in the area of the efficiency of parallel processing of point clouds and the quality of the research results often exceeds that of results produced using commercially available software tools.

ACKNOWLEDGEMENTS

The paper discusses the results obtained by a team of researchers at the Dephos Software Company, with headquarters in Zabierzów, near of-Kraków in Poland, as part of the following project financed by the European Union: "Research on large scale storage, sharing, and processing of spatial laser data."

REFERENCES

Będkowski J., Majek K., NuechterA., 2013. General Purpose Computing on Graphics Processing Units for Robotic Applications. *Journal of Software Engineering for Robotics*, 4(1), pp. 23-33.

Fischler M. A., Bolles R., 1980. Random sample consensus. A paradigm for model fitting with applications to image analysis and automated cartography, in: *Proc. 1980 Image Understanding Workshop* (College Park, Md., Apr 1980) L. S. Baurnann, Ed, Science Applications, McLean, Va., 1980, pp. 71–88.

Fischler M. A., Bolles R. C., 1987. Readings in computer vision: issues, problems, principles, and paradigms, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987, Ch. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, pp. 726–740.

Franke R, Nielson G., 1980. Smooth Interpolation of Large Sets of Scattered Data. *International Journal for Numerical Methods in Engineering*, 15(11), pp. 1691-1704.

Han S.H., Heo J., Sohn H., G., Yu K., 2009. Parallel Processing Method for Airborne Laser Scanning Data Using a PC Cluster and a Virtual Grid. *Sensors*, 9, pp. 2555-2573.

Pyka K., Rzonca A., 2006. Badanie jakości radiometrycznej ortofotogramów sporządzonych na drodze integracji fotogrametrii bliskiego zasięgu i skanowania laserowego, Archiwum Fotogrametrii, Kartografii i Teledetekcji, 16, pp. 515-526.

Rzonca A., 2006. Integracja danych skanerowych i fotogrametrycznych dla celów inwentaryzacji architektonicznej. *Geodezja*, 13(1), Kraków

Rzonca A., 2013. Integracja danych pozyskiwanych metodami fotogrametrycznymi i skanowania laserowego przy inwentaryzacji obiektów zabytkowych, Wydawnictwo AGH, Kraków 2013.

Zeng, X., He, W. (2009, October). GPGPU-based parallel processing of massive LiDAR point cloud. In Sixth International Symposium on Multispectral Image Processing and Pattern Recognition (pp. 749716-749716). International Society for Optics and Photonics.

http://www.asprs.org/a/society/committees/standards/asprs_las_format_v12.pdf

http://www.nvidia.pl/object/cuda-parallel-computing-pl.html

ZASTOSOWANIE OBLICZEŃ RÓWNOLEGŁYCH DO MASOWEGO PRZETWARZANIA DANYCH LASEROWYCH

Streszczenie

Publikacja ma na celu przedstawienie części wyników badań, jakie zrealizował zespół badawczy firmy Dephos Software w ramach projektu finansowanego przez UE pt. "Badania nad masowym przechowywaniem, udostępnianiem i przetwarzaniem przestrzennych danych laserowych". Na wstępie publikacji autorzy przedstawiają zasady organizacji algorytmu spełniającego wymogi obliczeń równoległych oraz przybliżają genezę pomysłu prowadzenia badań nad zastosowaniem procesorów graficznych do masowego przetwarzania danych skaningowych. Następnie autorzy prezentują wyniki oceny wydajności działania szeregu różnych procesów przetwarzania danych laserowych, które udało się zasadniczo przyspieszyć dzięki obliczeniom równoległym. Procesy te dzielą się na procesy podstawowe (generowanie ortoobrazów z chmur punktów, kolorowanie chmur punktów, transformacja, generowanie siatki regularnej) oraz procesy zaawansowane (wykrywanie płaszczyzn i krawędzi, klasyfikacja chmur punktów, analiza danych w celu kontroli jakości danych). W większości przypadków algorytmy musiały zostać opracowane całkowicie od nowa pod kątem wymogów przetwarzania równoległego, część korzysta z wcześniejszego dorobku technologicznego firmy Dephos Software, będąc dostosowana do równoległej metody obliczeń w ramach przeprowadzonych badań. W każdym z tych procesów określono czas działania dla typowej ilości danych przetwarzanych, co potwierdziło wysoką wydajność rozwiązań i sens zastosowania obliczeń równoległych w odniesieniu do danych skaningowych. Obliczenia równoległe dzięki swojej wysokiej wydajności otwierają nowe możliwości w tworzeniu i organizacji procesów przetwarzania danych pochodzących ze skaningu laserowego.

Dane autorów:

Janusz Będkowski e-mail: januszbedkowski@gmail.com

Ryszard Bratuś e-mail: rbratus@dephos.com

Marcin Prochaska e-mail: mprochaska@dephos.com

Dr inż. Antoni Rzonca e-mail: arz@agh.edu.pl

> Przesłano: 21.10.2015 Zaakceptowano: 15.12.2015